

Seminario Fuentes de datos Biomédicas y Web Semántica

Relación entre la Esperanza de Vida y el consumo de Agua

Urko Alli Barrena, Paula Gregorio Losada, Victoria Garcia Lovelle

2023-12-10

Contents

1. Introducción	2
2. Objetivos	2
3. Carga de librerías	2
4. Obtención de tablas	3
4.1. Tabla de Esperanza de Vida	3
4.2. Tabla de la Cantidad de Agua	5
4.3. Tabla de la Calidad del Agua	7
4.4. Tabla de Presupuesto para la Potabilización del Agua	9
5. Joins y Gráficas	10
5.1. Relación entre Esperanza de Vida y Cantidad	10
5.2. Relación entre Cantidad y Presupuesto	12
5.3. Relación entre Esperanza de Vida y Calidad	14
5.4. Relación entre Esperanza de Vida, Calidad, Presupuesto y Cantidad.	16
6. Conclusión	18
Anexo I	19
Referencias	21



1. Introducción

En el desarrollo de este seminario, analizaremos cómo diversos aspectos relacionados con la calidad y cantidad de agua consumida pueden influir en la esperanza de vida en las diferentes Comunidades Autónomas de España.

Exploraremos detalladamente cómo la calidad del agua, medida a través de diferentes parámetros, y la cantidad de agua, pueden ser factores determinantes para la salud y, consecuentemente, para la esperanza de vida de las Comunidades Autónomas.

2. Objetivos

El objetivo principal de este seminario es analizar de manera formal cómo diversos aspectos relacionados con la calidad y la cantidad de agua consumida pueden afectar en la esperanza de vida en las distintas Comunidades Autónomas de España. Se explorarán los siguientes puntos:

1. **Observar si la calidad del agua afecta directamente en la esperanza de vida:** Investigar cómo la calidad del agua puede tener repercusiones directas en la esperanza de vida de la población española por Comunidades Autónomas.
2. **Observar si la cantidad de agua consumida afecta directamente en la esperanza de vida:** Analizar de manera detallada cómo la cantidad de agua consumida puede influir en la esperanza de vida de la población española por Comunidades Autónomas.
3. **Observar si afecta el presupuesto para potabilizar el agua con la cantidad de agua consumida:** Evaluar la relación entre el presupuesto destinado a la potabilización del agua y la cantidad de agua consumida, explorando posibles implicaciones para la esperanza de vida.
4. **Impacto combinado de la calidad y cantidad de agua en la Esperanza de Vida:** Estudiar cómo la interacción entre la calidad y la cantidad de agua consumida puede tener un impacto significativo en la esperanza de vida de la población española por Comunidades Autónomas.

3. Carga de librerías

1. `pdftools`: Esta librería se utilizó para extraer datos de archivos PDF. En particular, se empleó para obtener información relevante sobre la calidad del agua de un informe en formato PDF.

```
library(pdftools)
```

2. `tidyverse`: Un conjunto de paquetes que facilitan la manipulación y visualización de datos en R. Incluye librerías como `tidyr` y `dplyr`, las cuales fueron esenciales para organizar y transformar datos.

```
library(tidyverse)
```

3. `tidyjson`: Esta librería fue útil para trabajar con datos en formato JSON. Fue esencial para analizar y extraer información de archivos JSON relacionados con la esperanza de vida y la cantidad de agua.

```
library(tidyjson)
```

4. rjson: Se utilizó para cargar y procesar datos almacenados en formato JSON. Facilita la manipulación de datos estructurados y su conversión a formatos compatibles con R.

```
library(rjson)
```

4. Obtención de tablas

4.1. Tabla de Esperanza de Vida

Para obtener la tabla de esperanza de vida se cargan los datos desde un archivo Json.

```
[{"COD":"IDB424788", "Nombre":"Andalucía. Mortalidad. Esperanza de vida. Total. 25 años. Anual. Dato base. Total. ", "T3_Unidad":"Años", "T3_Escala":" ",  
  "MetaData":[{"Id":8997, "Variable":{"Id":70, "Nombre":"Comunidades y Ciudades Autónomas", "Codigo":"1"}  
  ],  
  "Nombre":"Andalucía", "Codigo":"01"},  
  {"Id":12544, "Variable":{"Id":259, "Nombre":"Fecundidad, Nupcialidad, Mortalidad, Indicadores de población e Indicadores de migraciones", "Codigo":""}  
  },  
  "Nombre":"Mortalidad", "Codigo":""},  
  {"Id":284290, "Variable":{"Id":260, "Nombre":"Conceptos Demográficos", "Codigo":""}  
  },  
  "Nombre":"Esperanza de vida", "Codigo":""},  
  {"Id":451, "Variable":{"Id":18, "Nombre":"Sexo", "Codigo":""}  
  },  
  "Nombre":"Ambos sexos", "Codigo":"0"},  
  {"Id":15344, "Variable":{"Id":355, "Nombre":"Valores simples de edad", "Codigo":"15"},  
  "Nombre":"25 años", "Codigo":"025"},  
  {"Id":14954, "Variable":{"Id":310, "Nombre":"Periodicidad", "Codigo":""}  
  },  
  "Nombre":"Anual", "Codigo":""},  
  {"Id":72, "Variable":{"Id":3, "Nombre":"Tipo de dato", "Codigo":"7"},  
  "Nombre":"Dato base", "Codigo":"0"},  
  {"Id":21331, "Variable":{"Id":462, "Nombre":"TOTALES CLASIFICACION EDUCACIÓN", "Codigo":""}  
  },  
  "Nombre":"Total", "Codigo":""},  
  ],  
  "Data":[{"Fecha":"2021-01-01T00:00:00.000+01:00", "T3_TipoDato":"Definitivo", "T3_Periodo":"A", "Anyo":2021, "Valor":56.97}  
  ,{"Fecha":"2020-01-01T00:00:00.000+01:00", "T3_TipoDato":"Definitivo", "T3_Periodo":"A", "Anyo":2020, "Valor":57.01}  
  ,{"Fecha":"2019-01-01T00:00:00.000+01:00", "T3_TipoDato":"Definitivo", "T3_Periodo":"A", "Anyo":2019, "Valor":57.7}  
  ,{"Fecha":"2018-01-01T00:00:00.000+01:00", "T3_TipoDato":"Definitivo", "T3_Periodo":"A", "Anyo":2018, "Valor":57.25}  
  ,{"Fecha":"2017-01-01T00:00:00.000+01:00", "T3_TipoDato":"Definitivo", "T3_Periodo":"A", "Anyo":2017, "Valor":57.3}  
  ,{"Fecha":"2016-01-01T00:00:00.000+01:00", "T3_TipoDato":"Definitivo", "T3_Periodo":"A", "Anyo":2016, "Valor":57.3}  
  ]  
},  
{"COD":"IDB424787", "Nombre":"Andalucía. Mortalidad. Esperanza de vida. Total. 25 años. Anual. Dato base. Educación primaria e inferior. ", "T3  
_Unidad":"Años", "T3_Escala":" ", "MetaData":[{"Id":8997, "Variable":{"Id":70, "Nombre":"Comunidades y Ciudades Autónomas", "Codigo":"1"}  
  ],  
  "Nombre":"Andalucía", "Codigo":"01"},  
  {"Id":12544, "Variable":{"Id":259, "Nombre":"Fecundidad, Nupcialidad, Mortalidad, Indicadores de población e Indicadores de migraciones", "Codigo":""}  
  },  
  "Nombre":"Mortalidad", "Codigo":""},  
  {"Id":284290, "Variable":{"Id":260, "Nombre":"Conceptos Demográficos", "Codigo":""}  
  },  
  "Nombre":"Esperanza de vida", "Codigo":""}
```

Esto es una porción de todo el archivo JSON, del cual se tiene que seleccionar los datos deseados.

Una vez cargados los datos, se organizan en un formato de tabla y se realiza un análisis para saber qué tipos contiene la variable EsperanzaVida.

```
archivoJson <- fromJSON(file = "INPUT/data/EsperanzaVida.json")  
  
esperanzaVida <- spread_all(archivoJson)  
  
tibble1<-esperanzaVida %>%  
  gather_object %>%  
  json_types %>%  
  count(name, type)
```

Para obtener la tabla final se ha accedido a aquel que era de tipo array que contiene los valores de interés y organiza estos datos en una tabla.

```
arrayData<-esperanzaVida %>%  
  enter_object(Data) %>%  
  gather_array %>%  
  spread_all %>%  
  select(Nombre, Anyo, Valor)
```

El atributo Nombre contiene una cadena de caracteres muy larga separada por puntos. Entre toda la información están los nombres de las diferentes Comunidades Autónomas, por lo que se separa la cadena de texto y se obtienen únicamente estos nombres.

```
partes <- strsplit(arrayData$Nombre, "\\.")

comunidadesAutonomas<-c()
for (i in partes){
  comunidadesAutonomas<-c(comunidadesAutonomas,i[1])
}

arrayData$Nombre<-comunidadesAutonomas
```

De la tabla se obtiene el año 2020 de cada Comunidad Autónoma y se obtiene como valor la media de la esperanza de vida en ese año.

Para tener todas las tablas con el mismo estilo, se han pasado los nombres de las comunidades autónomas a mayúsculas.

```
tablaComunidadesAñoValor<- as_tibble(arrayData)
attr(tablaComunidadesAñoValor, "JSON") <- NULL

tablaEsperanzaDeVida <- tablaComunidadesAñoValor %>%
  filter(Anyo==2020)%>%
  group_by(Anyo, Nombre) %>%
  summarize(EsperanzaDeVida = mean(Valor, na.rm = TRUE))%>%
  rename(Año=Anyo, ComunidadAutonoma=Nombre)

tablaEsperanzaDeVidaFinal <- tablaEsperanzaDeVida %>%
  mutate(ComunidadAutonoma = toupper(ComunidadAutonoma))

colnames(tablaEsperanzaDeVidaFinal) <- c("Anio", "ComunidadAutonoma", "EsperanzaDeVida")
```

En la tabla final, se puede observar la media de la esperanza de vida de cada Comunidad Autónoma en el año 2020.

```
tablaEsperanzaDeVidaFinal
```

```
## # A tibble: 19 x 3
## # Groups:   Anio [1]
##   Anio ComunidadAutonoma      EsperanzaDeVida
##   <dbl> <chr>                <dbl>
## 1 2020 ANDALUCÍA            28.0
## 2 2020 ARAGÓN              28.4
## 3 2020 ASTURIAS, PRINCIPADO DE 28.1
## 4 2020 BALEARS, ILLES      29.3
## 5 2020 CANARIAS            29.1
## 6 2020 CANTABRIA           28.9
## 7 2020 CASTILLA - LA MANCHA  27.5
## 8 2020 CASTILLA Y LEÓN     28.5
## 9 2020 CATALUÑA            28.3
## 10 2020 CEUTA              26.7
## 11 2020 COMUNITAT VALENCIANA 28.5
```

## 12	2020	EXTREMADURA	27.9
## 13	2020	GALICIA	29.3
## 14	2020	MADRID, COMUNIDAD DE	28.0
## 15	2020	MELILLA	26.3
## 16	2020	MURCIA, REGIÓN DE	28.4
## 17	2020	NAVARRA, COMUNIDAD FORAL DE	29.0
## 18	2020	PAÍS VASCO	28.8
## 19	2020	RIOJA, LA	28.6

4.2. Tabla de la Cantidad de Agua

Para obtener la tabla de cantidad de agua se cargan los datos desde un archivo Json.

```
[{"Nombre":"España, Volumen total de agua registrada y distribuida por tipo de usuario", "MetaData":[{"Variable":{"Id":349, "Nombre":"Total Nacional", "Codigo":"NAC"}}, {"Nombre":"España", "Codigo":"00"}], {"Variable":{"Nombre":"grupos de usuarios e importe", "Codigo":"gruposdeusuarioseimporte"}}, {"Nombre":"Volumen total de agua registrada y distribuida por tipo de usuario", "Codigo":"volumentotaldeaguaregistradaydistribuidaportipodeusuario"}], "Data":[{"NombrePeriodo":"2020", "Valor":3177815.0}, {"NombrePeriodo":"2018", "Valor":3188055.0}, {"NombrePeriodo":"2016", "Valor":3199910.0}, {"NombrePeriodo":"2014", "Valor":3214034.0}, {"NombrePeriodo":"2013", "Valor":3210843.0}, {"NombrePeriodo":"2012", "Valor":3338028.0}, {"NombrePeriodo":"2011", "Valor":3381318.0}, {"NombrePeriodo":"2010", "Valor":3393271.0}, {"NombrePeriodo":"2009", "Valor":3500578.0}, {"NombrePeriodo":"2008", "Valor":3731399.0}, {"NombrePeriodo":"2007", "Valor":3778036.0}, {"NombrePeriodo":"2006", "Valor":3913059.0}, {"NombrePeriodo":"2005", "Valor":4002180.0}, {"NombrePeriodo":"2004", "Valor":4042399.0}, {"NombrePeriodo":"2003", "Valor":4019615.0}, {"NombrePeriodo":"2002", "Valor":3855697.0}, {"NombrePeriodo":"2001", "Valor":3870650.0}, {"NombrePeriodo":"2000", "Valor":3781680.0}], [{"Nombre":"España, - Sectores económicos", "MetaData":[{"Variable":{"Id":349, "Nombre":"Total Nacional", "Codigo":"NAC"}}, {"Nombre":"España", "Codigo":"00"}], {"Variable":{"Nombre":"grupos de usuarios e importe", "Codigo":"gruposdeusuarioseimporte"}}, {"Nombre":"- Sectores económicos", "Codigo":"sectoreseconomicos"}]
```

Esto es una porción de todo el archivo JSON, del cual se tiene que seleccionar los datos deseados.

Una vez cargados los datos, se organizan en un formato de tabla y se realiza un análisis para saber qué tipos contiene la variable cantidadAgua.

```
archivoJsonCantidad <- fromJSON(file = "INPUT/data/CantidadAgua.json")

cantidadAgua <- spread_all(archivoJsonCantidad)

cantidadAgua%>%
  gather_object%>%
  json_types%>%
  count(name,type)
```

```
## # A tibble: 3 x 3
##   name      type      n
##   <chr>    <fct> <int>
## 1 Data      array    133
## 2 MetaData array    133
## 3 Nombre   string   133
```

Para poder obtener la tabla final de la cantidad de agua consumida en cada Comunidad Autónoma, se ha accedido al atributo de tipo array que contiene la información de interés que es Data y organiza estos datos en una tabla.

Seguidamente, como únicamente necesitamos las comunidades autónomas y están escritas en una larga cadena de texto, dividimos la cadena de texto y obtenemos de ella el nombre de las Comunidades Autónomas.

Esa nueva variable con las Comunidades Autónomas se añaden a la tabla final y se elimina donde ponga España.

```
arrayDataCantidad<-cantidadAgua%>%
  enter_object(Data)%>%
  gather_array%>%
  spread_all%>%
  select(-document.id,-array.index)

cadenas <- strsplit(arrayDataCantidad$Nombre, "\\,")

comunidadesAutonomasCantidad<-c()
for (i in cadenas){
  comunidadesAutonomasCantidad<-c(comunidadesAutonomasCantidad,i[1])
}

arrayDataCantidad$Nombre<-comunidadesAutonomasCantidad

tabla <- arrayDataCantidad %>%
  filter(!(Nombre == "España")) %>%
  select(Nombre, NombrePeriodo, Valor)
```

Se filtra únicamente por el año 2020, que es el año de interés, se realiza una media de los valores obtenidos en el año 2020 de cada comunidad autónoma.

Para tener todas las tablas con el mismo estilo, se han pasado los nombres de las Comunidades Autónomas a mayúsculas y se han renombrado.

Por último se ha pasado a tipo entero la columna años.

```
tablaCantidadDeAgua <- tabla %>%
  filter(NombrePeriodo ==2020)%>%
  group_by(NombrePeriodo , Nombre) %>%
  summarize(Cantidad = mean(Valor, na.rm = TRUE))%>%
  rename(Año=NombrePeriodo,ComunidadAutonoma=Nombre)

tablaCantidadDeAgua <- tablaCantidadDeAgua %>%
  mutate(ComunidadAutonoma = toupper(ComunidadAutonoma))

tablaCantidadDeAgua[4,2] <- c("BALEARNS, ILLES")
tablaCantidadDeAgua[8,2] <- c("CASTILLA - LA MANCHA")
tablaCantidadDeAgua[14,2] <- c("MADRID, COMUNIDAD DE")
tablaCantidadDeAgua[15,2] <- c("MURCIA, REGIÓN DE")
tablaCantidadDeAgua[16,2] <- c("NAVARRA, COMUNIDAD FORAL DE")
tablaCantidadDeAgua[18,2] <- c("RIOJA, LA")

colnames(tablaCantidadDeAgua) <- c("Anio", "ComunidadAutonoma", "Cantidad")

tablaCantidadDeAguaFinal <- tablaCantidadDeAgua %>%
  mutate_at(vars(Anio), as.integer)
```

En la tabla final de la cantidad de agua, se puede observar la cantidad de agua consumida por cada Comunidad Autónoma en el año 2020.

```
tablaCantidadDeAguaFinal
```

```
## # A tibble: 18 x 3
## # Groups:   Anio [1]
##   Anio ComunidadAutonoma      Cantidad
##   <int> <chr>              <dbl>
## 1 2020 ANDALUCÍA          262867
## 2 2020 ARAGÓN             38689.
## 3 2020 ASTURIAS           28342.
## 4 2020 BALEARS, ILLES     42033.
## 5 2020 CANARIAS           79923.
## 6 2020 CANTABRIA          20367.
## 7 2020 CASTILLA Y LEÓN     73272.
## 8 2020 CASTILLA - LA MANCHA 60920.
## 9 2020 CATALUÑA          239767.
## 10 2020 CEUTA Y MELILLA      5507.
## 11 2020 COMUNITAT VALENCIANA 188854.
## 12 2020 EXTREMADURA        28064.
## 13 2020 GALICIA            71149
## 14 2020 MADRID, COMUNIDAD DE 216232.
## 15 2020 MURCIA, REGIÓN DE    57000.
## 16 2020 NAVARRA, COMUNIDAD FORAL DE 22551.
## 17 2020 PAÍS VASCO          52635.
## 18 2020 RIOJA, LA           9448.
```

4.3. Tabla de la Calidad del Agua

Este código extrae y procesa información relevante sobre la calidad del agua de un informe en PDF. El PDF tiene 217 páginas de las cuales la página deseada es la 17 que contiene la tabla que queremos obtener. La tabla en el PDF se muestra así:

Comunidad Autónoma	Num. de Municipios	Zonas de baño	Puntos de muestreo	Aguas "2"	Aguas "1"	Aguas "0"	Aguas SCF
ANDALUCÍA	58	63	70	3	36	27	4
ARAGÓN	11	11	12	3	7	1	1
ASTURIAS	1	1	1	0	0	1	0
CASTILLA-LA MANCHA	28	39	43	24	7	12	0
CASTILLA Y LEÓN	2	2	2	0	1	1	0
CATALUÑA	9	10	11	3	8	0	0
EXTREMADURA	17	17	17	0	0	0	17
GALICIA	53	54	68	10	45	13	0
MADRID	6	6	7	0	2	5	0
MURCIA	3	3	3	0	0	3	0
NAVARRA	11	11	11	4	5	2	0
RIOJA	1	1	1	0	1	0	0
VALENCIA	2	2	2	0	1	1	0
TOTAL	202	220	248	48	112	66	22

Tabla 52. Distribución territorial y calidad sanitaria de los puntos de muestreo de las zonas de baño continentales

Para poder obtener esta tabla y poder utilizarla para el estudio, primero se convierte el contenido del PDF a texto y luego selecciona un rango específico de líneas en una página particular del documento (página 17), con el objetivo de obtener datos clave relacionados con la calidad del agua.

```

ruta_pdf <- pdf_text("INPUT/data/report_Cap.3_part2._Libro_blanco_del_agua.pdf")
pagina <- ruta_pdf[17]
lineas <- strsplit(pagina, "\n")[[1]]
linea_deseada <- lineas[5:21]

```

En este código se dividen las líneas deseadas usando espacios como delimitadores. Luego, crea dos conjuntos de datos: `datos_obtenidos`, excluyendo las primeras tres dimensiones no necesarias, y `datosDeInteres`, que contiene la información esencial sobre las comunidades autónomas y los valores de la calidad del agua.

Además, se añaden manualmente filas específicas que no se extrajeron correctamente del informe.

```

datos_divididos <- strsplit(linea_deseada, "\\s+")
datos_obtenidos <- datos_divididos[-c(1:3)]

datosDeInteres <- datos_obtenidos[-c(length(datos_obtenidos))]
datosDeInteres[[4]] <- c("CASTILLA - LA MANCHA", "28", "39", "43", "24", "7", "12", "0")
datosDeInteres[[5]] <- c("CASTILLA Y LEÓN", "2", "2", "2", "0", "1", "1", "0")
datosDeInteres[[9]] <- c("MADRID, COMUNIDAD DE", "6", "6", "7", "0", "2", "5", "0")
datosDeInteres[[10]] <- c("MURCIA, REGIÓN DE", "3", "3", "3", "0", "0", "3", "0")
datosDeInteres[[11]] <- c("NAVARRA, COMUNIDAD FORAL DE", "11", "11", "11", "4", "5", "2", "0")
datosDeInteres[[12]] <- c("RIOJA, LA", "1", "1", "1", "0", "1", "0", "0")
datosDeInteres[[13]] <- c("COMUNITAT VALENCIANA", "2", "2", "2", "0", "1", "1", "0")

```

En cada iteración del bucle, se agrega la fila representada por el elemento `i` al data frame `tablaCalidadDeAgua` utilizando la función `rbind()`.

Por último se renombra y se cambia a tipo numérico a aquellas columnas que sea necesario.

```

tablaCalidadDeAgua<-data.frame()
for (i in datosDeInteres){
  tablaCalidadDeAgua<-rbind(tablaCalidadDeAgua,i)
}

colnames(tablaCalidadDeAgua) <- c("ComunidadAutonoma", "NumdeMunicipios", "ZonasdeBaño", "PuntosdeMuestreo", "Aguas2", "Aguas1", "Aguas0")

colNumericas <- c("NumdeMunicipios", "ZonasdeBaño", "PuntosdeMuestreo", "Aguas2", "Aguas1", "Aguas0")
tablaCalidadDeAguaFinal <- tablaCalidadDeAgua %>%
  mutate_at(vars(colNumericas), as.integer) %>%
  select(ComunidadAutonoma, Aguas2, Aguas1, Aguas0)

```

```

tablaCalidadDeAguaFinal

```

##	ComunidadAutonoma	Aguas2	Aguas1	Aguas0
## 1	ANDALUCÍA	3	36	27
## 2	ARAGÓN	3	7	1
## 3	ASTURIAS	0	0	1
## 4	CASTILLA - LA MANCHA	24	7	12
## 5	CASTILLA Y LEÓN	0	1	1
## 6	CATALUÑA	3	8	0
## 7	EXTREMADURA	0	0	0
## 8	GALICIA	10	45	13
## 9	MADRID, COMUNIDAD DE	0	2	5
## 10	MURCIA, REGIÓN DE	0	0	3

## 11	NAVARRA, COMUNIDAD FORAL DE	4	5	2
## 12	RIOJA, LA	0	1	0
## 13	COMUNITAT VALENCIANA	0	1	1

En esta tabla final se pueden observar unas columnas llamadas Aguas1, Aguas2 y Aguas0, vamos a ver a qué se refieren:

- Aguas2: Aguas aptas, de muy buena calidad.
- Aguas1: Aguas aptas, de buena calidad.
- Aguas0: Aguas no aptas, mala calidad.

Están evaluados diferentes puntos de muestreo en cada Comunidad Autónoma, de los cuales se puede observar la calidad general de cada comunidad en concreto.

4.4. Tabla de Presupuesto para la Potabilización del Agua

Para obtener esta tabla se ha importado un archivo CSV utilizando la función `read_csv` y después se renombran las columnas para tener unificados todos los nombres de las diferentes tablas.

```
Total Nacional,"Comunidades y Ciudades Aut  nomas","Grupos de usuarios e importe","periodo","Total"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2020,"3.177.815"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2018,"3.188.055"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2016,"3.199.910"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2014,"3.214.034"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2013,"3.210.843"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2012,"3.338.028"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2011,"3.381.318"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2010,"3.393.271"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2009,"3.500.578"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2008,"3.731.399"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2007,"3.778.036"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2006,"3.913.059"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2005,"4.002.180"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2004,"4.042.399"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2003,"4.019.615"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2002,"3.855.697"
Total Nacional,NA,"Volumen total de agua registrada y distribuida por tipo de usuario",2001,"3.870.650"
```

Esto es una porción de todo el archivo CSV, del cual se tiene que seleccionar los datos deseados.

Una vez cargados los datos, se renombran las columnas para poder tener unificados todos los nombres.

```
summodificado <- read_csv("INPUT/data/summodificado.csv")
colnames(summodificado) <- c("TotalNacional", "ComunidadAutonoma", "GruposDeUsuarioEImporte", "Anio", "P
```

Se seleccionan las filas donde la columna “GruposDeUsuarioEImporte” es igual a “Importe total de la inversión en los servicios de suministro” y la columna “Anio” es igual a “2020”.

Delante de cada nombre de la Comunidad Autónoma existe un número por lo que se elimina y se convierten los nombres de las comunidades a mayúsculas.

Se eliminan los puntos de la columna presupuesto y se pasa a tipo entero. Y por último se descarta lo que no es necesario.

```

tablaPresupuestos <- summodificado%>%
  filter(GruposDeUsuarioEImporte=="Importe total de la inversión en los servicios de suministro" & Anio
  mutate(ComunidadAutonoma = gsub("^\\d+\\s*", "", ComunidadAutonoma)) %>%
  mutate(ComunidadAutonoma = toupper(ComunidadAutonoma)) %>%
  select (.data = ., Anio, ComunidadAutonoma:Presupuesto) %>%
  drop_na()

tablaPresupuestos$Presupuesto <- gsub("\\.", "", tablaPresupuestos$Presupuesto)
tablaPresupuestos$Presupuesto <- as.integer(tablaPresupuestos$Presupuesto)
tablaPresupuestosFinal <- tablaPresupuestos[,-3]

```

Así obtenemos los datos de los presupuesto para la potabilización del agua en el año 2020 en las diferentes Comunidades Autónomas de España

```
tablaPresupuestosFinal
```

```
## # A tibble: 17 x 3
##   Anio ComunidadAutonoma      Presupuesto
##   <dbl> <chr>              <int>
## 1 2020 ANDALUCÍA          43361
## 2 2020 ARAGÓN             691
## 3 2020 ASTURIAS, PRINCIPADO DE 757
## 4 2020 BALEARS, ILLES    14514
## 5 2020 CANARIAS          4015
## 6 2020 CANTABRIA         223
## 7 2020 CASTILLA Y LEÓN    8083
## 8 2020 CASTILLA - LA MANCHA 2397
## 9 2020 CATALUÑA          30222
## 10 2020 COMUNITAT VALENCIANA 39274
## 11 2020 EXTREMADURA       2263
## 12 2020 GALICIA           2594
## 13 2020 MADRID, COMUNIDAD DE 80532
## 14 2020 MURCIA, REGIÓN DE   3170
## 15 2020 NAVARRA, COMUNIDAD FORAL DE 5704
## 16 2020 PAÍS VASCO        3001
## 17 2020 RIOJA, LA         304

```

5. Joins y Gráficas

En este apartado realizamos una cantidad de cuatro joins, los cuales explicaremos a continuación.

5.1. Relación entre Esperanza de Vida y Cantidad

```

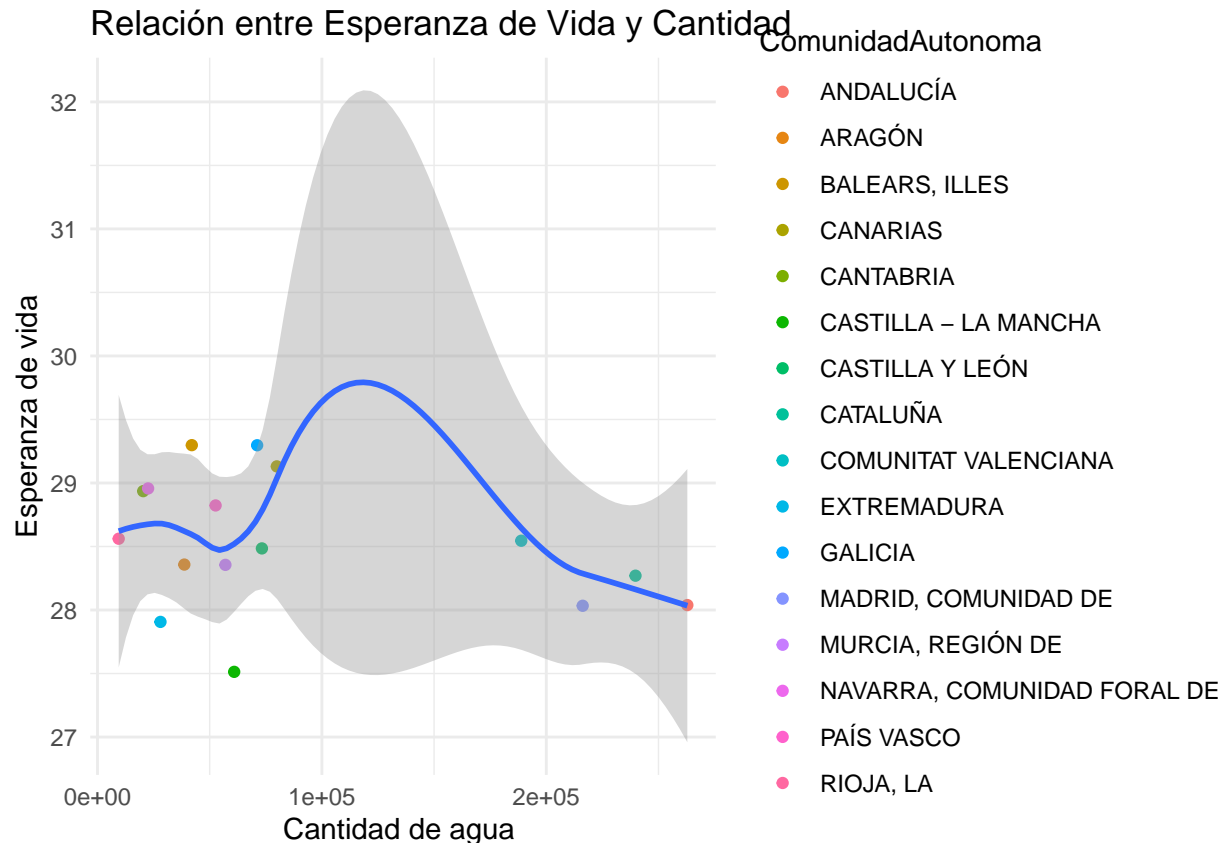
EsperanzayCantidad<- tablaEsperanzaDeVidaFinal%>%
  left_join(x=., y=tablaCantidadDeAguaFinal, by=c("Anio","ComunidadAutonoma"))%>%
  group_by(ComunidadAutonoma) %>%
  drop_na()
EsperanzayCantidad

```

```
## # A tibble: 16 x 4
## # Groups:   ComunidadAutonoma [16]
##   Anio ComunidadAutonoma EsperanzaDeVida Cantidad
##   <dbl> <chr>                <dbl>    <dbl>
## 1 2020 ANDALUCÍA           28.0  262867
## 2 2020 ARAGÓN             28.4   38689.
## 3 2020 BALEARS, ILLES     29.3   42033.
## 4 2020 CANARIAS           29.1   79923.
## 5 2020 CANTABRIA          28.9   20367.
## 6 2020 CASTILLA - LA MANCHA 27.5   60920.
## 7 2020 CASTILLA Y LEÓN    28.5   73272.
## 8 2020 CATALUÑA           28.3  239767.
## 9 2020 COMUNITAT VALENCIANA 28.5  188854.
## 10 2020 EXTREMADURA       27.9   28064.
## 11 2020 GALICIA            29.3   71149
## 12 2020 MADRID, COMUNIDAD DE 28.0  216232.
## 13 2020 MURCIA, REGIÓN DE   28.4   57000.
## 14 2020 NAVARRA, COMUNIDAD FORAL DE 29.0   22551.
## 15 2020 PAÍS VASCO         28.8   52635.
## 16 2020 RIOJA, LA          28.6    9448.
```

En el fragmento de código anterior, combinamos mediante un *left_join()* las tablas de “tablaEsperanzaDeVidaFinal” y “tablaCantidadDeAguaFinal” usando sus claves comunes, las columnas “Anio” y “ComunidadAutonoma” ordenado en función de las Comunidades Autónomas.

```
grafEsperanzaCantidad <- ggplot(data=EsperanzayCantidad, aes(x=Cantidad, y=EsperanzaDeVida))+
  geom_point(aes(color=ComunidadAutonoma))+
  geom_smooth()+
  labs(title="Relación entre Esperanza de Vida y Cantidad",
        x="Cantidad de agua ",
        y="Esperanza de vida")+
  theme_minimal()
grafEsperanzaCantidad
```



Como podemos observar en esta caja de código, creamos la gráfica de la nueva tabla “EsperanzayCantidad”, creada a partir del `left_join()` anterior. En el eje de las abscisas se representa la cantidad de agua consumida, mientras que en el eje de las ordenadas encontramos la esperanza de vida. La leyenda muestra que cada comunidad es representada en la gráfica mediante un color del gradiente y en forma circular gracias a la función `geom_point(aes(color=ComunidadAutonoma))`. Con la función `geom_smooth()` creamos la línea de tendencia, ya que tenemos un gráfico de dispersión y así podemos entender el patrón que estamos estudiando con esta gráfica.

Nuestro principal objetivo es determinar la esperanza de vida. En esta tabla relacionamos la esperanza de vida con la cantidad de agua que se consume en cada Comunidad Autónoma. Podemos observar que no hay una relación directamente proporcional entre la cantidad de agua consumida y la esperanza de vida de cada Comunidad Autónoma. Comprobamos en el Anexo I que las Islas Baleares se encuentra en la undécima posición en cuanto al consumo de agua (42033.429) y en primer lugar en cuanto a la esperanza de vida (29,29776). Por otro lado, Andalucía está situada en el primer puesto en cuanto a la cantidad de agua que consumen (262867.000) y en el decimotercero en lo que respecta a la esperanza de vida (28.03881). Sin embargo, al observar la línea de tendencia de nuestro gráfico, notamos que los extremos no son lo ideal. Por lo tanto, la tendencia nos sugiere que aquellos ciudadanos con una esperanza de vida potencialmente alta son aquellos que consumen una cantidad intermedia de agua.

5.2. Relación entre Cantidad y Presupuesto

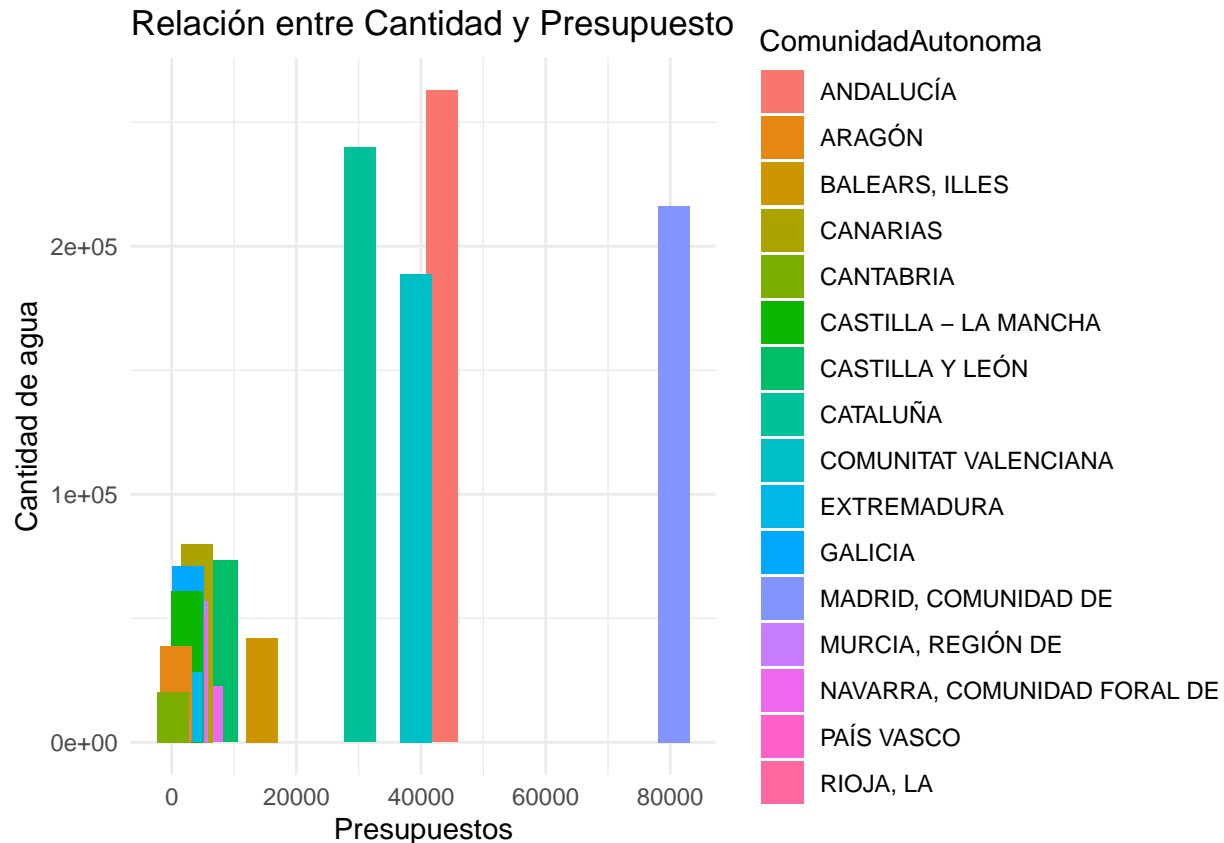
```
CantidadyPresupuesto<- tablaCantidadDeAguaFinal%>%
  left_join(x=., y=tablaPresupuestosFinal, by=c("Anio","ComunidadAutonoma")) %>%
  arrange(desc(Presupuesto)) %>%
```

```
drop_na()
CantidadPresupuesto
```

```
## # A tibble: 16 x 4
## # Groups:   Anio [1]
##   Anio ComunidadAutonoma      Cantidad Presupuesto
##   <dbl> <chr>                <dbl>         <int>
## 1 2020 MADRID, COMUNIDAD DE    216232.        80532
## 2 2020 ANDALUCÍA              262867         43361
## 3 2020 COMUNITAT VALENCIANA    188854.        39274
## 4 2020 CATALUÑA               239767.        30222
## 5 2020 BALEARs, ILLES         42033.         14514
## 6 2020 CASTILLA Y LEÓN        73272.          8083
## 7 2020 NAVARRA, COMUNIDAD FORAL DE 22551.         5704
## 8 2020 CANARIAS               79923.         4015
## 9 2020 MURCIA, REGIÓN DE       57000.         3170
## 10 2020 PAÍS VASCO             52635.         3001
## 11 2020 GALICIA                71149         2594
## 12 2020 CASTILLA - LA MANCHA    60920.         2397
## 13 2020 EXTREMADURA           28064.         2263
## 14 2020 ARAGÓN                 38689.          691
## 15 2020 RIOJA, LA              9448.          304
## 16 2020 CANTABRIA             20367.          223
```

En el fragmento de código anterior, combinamos a través de un *left_join()* las tablas de “tablaCantidad-DeAguaFinal” y “tablaPresupuestosFinal” gracias a sus claves comunes “Anio” y “ComunidadAutonoma”, ordenadas en función de las del presupuesto de manera descendente.

```
grafCantidadPresupuesto <- ggplot(data=CantidadPresupuesto, aes(x= Presupuesto, y= Cantidad, fill=Comun
  geom_bar(stat= "identity")+
  labs(title="Relación entre Cantidad y Presupuesto",
        x="Presupuestos",
        y="Cantidad de agua")+
  geom_bar(stat= "identity", width = 5000)+
  theme_minimal()
grafCantidadPresupuesto
```



Aquí creamos la gráfica de la nueva tabla “grafCantidadPresupuesto”, creada a partir del `left_join()` anterior. En el eje de las ordenadas encontramos la cantidad de agua consumida, y en el eje de las abscisas encontramos los presupuestos. Como podemos comprobar en la leyenda, gracias a la función `(...)fill=ComunidadAutonoma))+ geom_bar(stat= “identity”)`, cada comunidad será representada en la gráfica por un color perteneciente al gradiente y en forma de barra.

Al observar la tabla y el gráfico, es evidente que en este caso sí podemos apreciar una correlación significativa entre el presupuesto invertido en agua y su consumo. Lugares como la Comunidad de Madrid, Andalucía o la Comunidad Valenciana encabezan los presupuestos más altos y el consumo de agua mayor (puede comprobarse en el Anexo I). Por otro lado, La Rioja , Cantabria, Extremadura y Aragón, rondan los últimos puestos tanto en presupuesto como en cantidad. Podemos destacar Castilla-La Mancha, que con un gasto relativamente bajo en agua, tienen un consumo considerable, justo lo contrario de la Comunidad Foral de Navarra. Aun teniendo en cuenta las pocas comunidades que no se rigen por la normalidad de este gráfico, podemos afirmar que una mayor inversión por parte de las Comunidades Autónomas fomentan un mayor consumo de agua en sus habitantes.

5.3. Relación entre Esperanza de Vida y Calidad

```

EsperanzayCalidad<- tablaEsperanzaDeVidaFinal%>%
  left_join(x=., y=tablaCalidadDeAguaFinal, by=c("ComunidadAutonoma"))%>%
  group_by(ComunidadAutonoma) %>%
  drop_na()
EsperanzayCalidadFinal <- pivot_longer(data = EsperanzayCalidad, names_to = "CalidadAgua", values_to = 
EsperanzayCalidadFinal

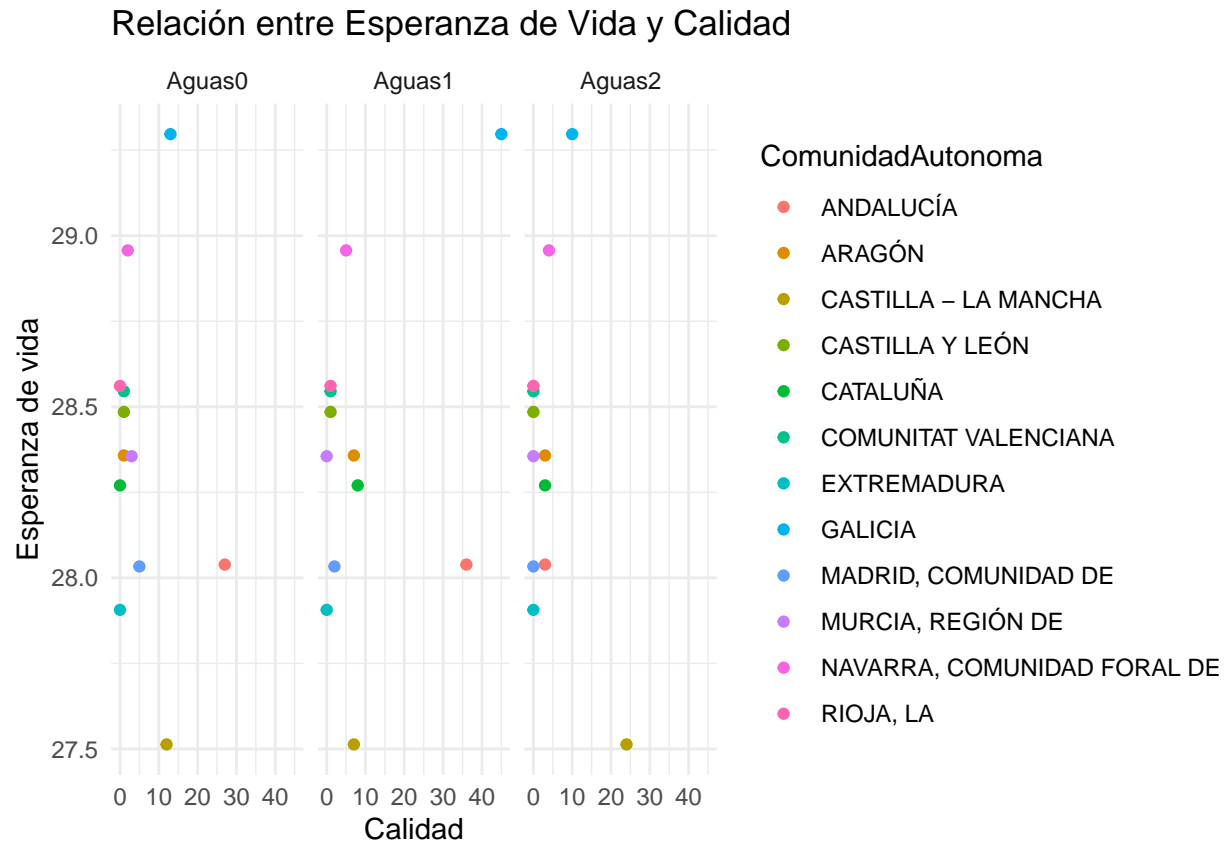
```

```
## # A tibble: 36 x 5
## # Groups:   ComunidadAutonoma [12]
##   Anio ComunidadAutonoma EsperanzaDeVida CalidadAgua ValoresCalidadAgua
##   <dbl> <chr>                <dbl> <chr>                <int>
## 1 2020 ANDALUCÍA                28.0 Aguas2                 3
## 2 2020 ANDALUCÍA                28.0 Aguas1                36
## 3 2020 ANDALUCÍA                28.0 Aguas0                27
## 4 2020 ARAGÓN                  28.4 Aguas2                 3
## 5 2020 ARAGÓN                  28.4 Aguas1                 7
## 6 2020 ARAGÓN                  28.4 Aguas0                 1
## 7 2020 CASTILLA - LA MANCHA    27.5 Aguas2                24
## 8 2020 CASTILLA - LA MANCHA    27.5 Aguas1                 7
## 9 2020 CASTILLA - LA MANCHA    27.5 Aguas0                12
## 10 2020 CASTILLA Y LEÓN        28.5 Aguas2                 0
## # i 26 more rows
```

En este código, combinamos con un *left_join()* las tablas de “tablaEsperanzaDeVidaFinal” y “tablaCalidadDeAguaFinal” mediante su clave común, la columna “ComunidadAutonoma”, ordenado en función de la Comunidad Autónoma. De esta tabla, consideraremos las columnas “Aguas2”, “Aguas1”, “Aguas 0”, las cuales son relevantes en nuestro estudio.

Gracias a un *pivot_longer()*, podemos crear una nueva columna “CalidadAgua”, agrupando las tres últimas columnas en distintas filas, y otra columna “ValoresCalidadAgua” con los valores correspondientes a esas variables, haciendo más fácil su posterior estudio en la gráfica.

```
grafEsperanzaCalidad <- ggplot(data = EsperanzayCalidadFinal, aes(x = ValoresCalidadAgua, y = EsperanzaDeVidaFinal)) +
  geom_point(aes(colour = ComunidadAutonoma)) +
  facet_wrap(facets = vars(CalidadAgua), nrow = 1)+
  labs(title="Relación entre Esperanza de Vida y Calidad",
        x="Calidad",
        y="Esperanza de vida")+
  theme_minimal()
grafEsperanzaCalidad
```



Para crear este gráfico a partir de la nueva tabla “CalidadyEsperanza”, utilizamos como eje de abscisas la variable “ValoresCalidadAgua”, mientras que de eje de ordenadas se encuentra la variable “EsperanzaDeVida”. Por medio de la función `geom_point()`, agregamos puntos al gráfico asignando colores en función de la Comunidad Autónoma que represente. Al tener varios datos de la calidad del agua, dividimos las facetas con la ayuda de la función `facet_wrap()` organizándolas en una sola fila.

Como se explica en el Apartado 3 de la obtención de tablas, los tres tipos de agua nos indicarán el grado de calidad de estas. Según los porcentajes obtenidos en el AnexoI, la Comunidad Autónoma que utiliza como fuente principal las aguas de tipo “Aguas 2” es Castilla La Mancha, pero su esperanza de vida es baja. Luego, Comunidades Autónomas como Murcia y Asturias utilizan el 100% de “Aguas 0” y aun así tienen una esperanza de vida en la media de lo normal. Por último, Galicia, líder en esperanza de vida, usa en su mayoría aguas de tipo “Aguas 1”. Concluyendo así que no hay ninguna relación a primera vista sobre la calidad de agua utilizada y la esperanza de vida.

5.4. Relación entre Esperanza de Vida, Calidad, Presupuesto y Cantidad.

```
tablaend<- EsperanzayCantidad %>%
  left_join(x=., y=CantidadyPresupuesto, by=c("Cantidad","ComunidadAutonoma","Anio")) %>%
  left_join(x=., y=EsperanzayCalidadFinal, by=c("ComunidadAutonoma")) %>%
  mutate(EsperanzaDeVida = coalesce(EsperanzaDeVida.x, EsperanzaDeVida.y)) %>%
  select(-EsperanzaDeVida.x, -EsperanzaDeVida.y)%>%
  mutate(Anio = coalesce(Anio.x, Anio.y))%>%
  select(-Anio.x, -Anio.y)%>%
  select(Anio, ComunidadAutonoma, everything())%>%
```

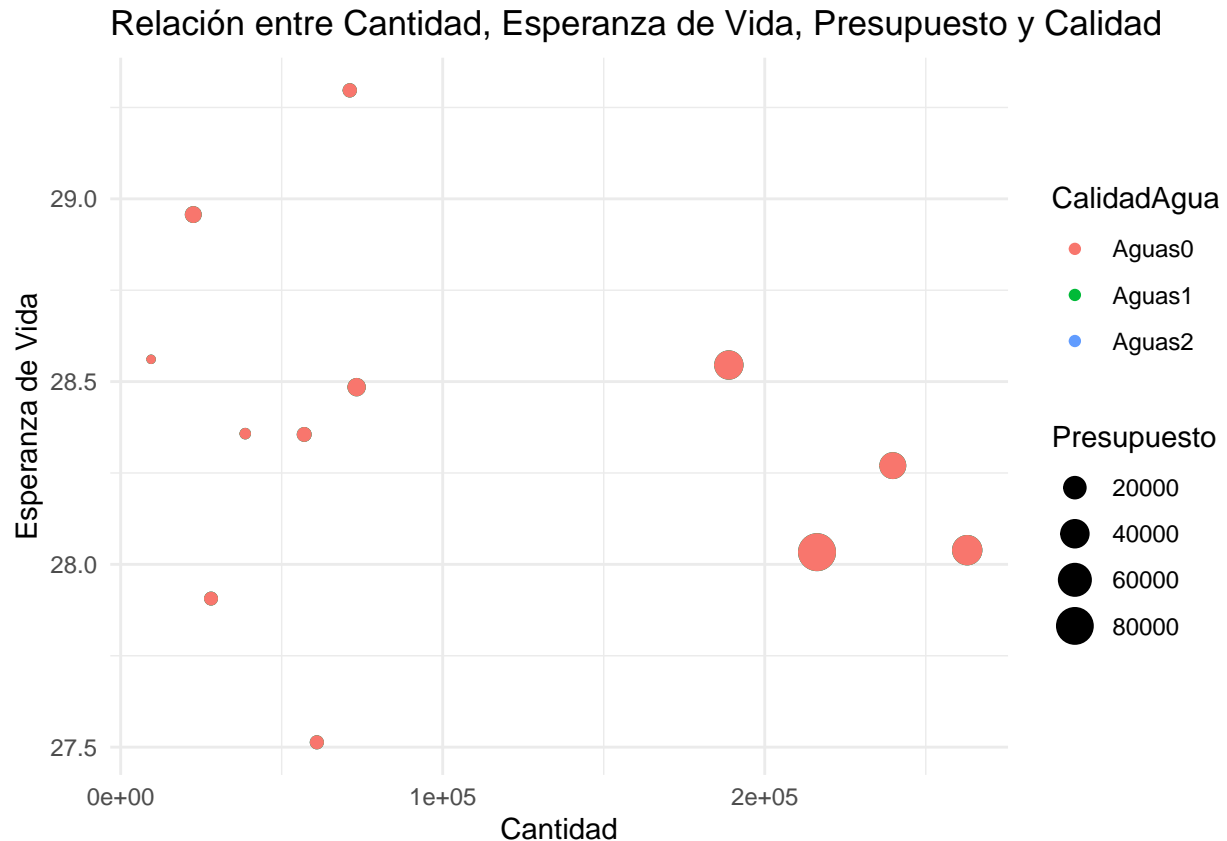


```
drop_na()
tablaend
```

```
## # A tibble: 36 x 7
## # Groups:   ComunidadAutonoma [12]
##   Anio ComunidadAutonoma Cantidad Presupuesto CalidadAgua ValoresCalidadAgua
##   <dbl> <chr>           <dbl>      <int> <chr>           <int>
## 1 2020 ANDALUCÍA      262867    43361 Aguas2           3
## 2 2020 ANDALUCÍA      262867    43361 Aguas1          36
## 3 2020 ANDALUCÍA      262867    43361 Aguas0          27
## 4 2020 ARAGÓN        38689.     691 Aguas2           3
## 5 2020 ARAGÓN        38689.     691 Aguas1           7
## 6 2020 ARAGÓN        38689.     691 Aguas0           1
## 7 2020 CASTILLA - LA MANC~ 60920.    2397 Aguas2          24
## 8 2020 CASTILLA - LA MANC~ 60920.    2397 Aguas1           7
## 9 2020 CASTILLA - LA MANC~ 60920.    2397 Aguas0          12
## 10 2020 CASTILLA Y LEÓN    73272.    8083 Aguas2           0
## # i 26 more rows
## # i 1 more variable: EsperanzaDeVida <dbl>
```

En este código, unimos mediante un *left_join()* las tablas de “EsperanzayCantidad” y “CantidadyPresupuesto” teniendo como clave común las columnas “ComunidadAutonoma”, “Cantidad” y “Anio”. Posteriormente, realizamos otro *left_join()* combinando las tablas “EsperanzayCantidad” junto con “EsperanzayCalidadFinal”, las cuales se unen gracias a la columna “ComunidadAutonoma”. A partir del resultado anterior, eliminaremos las columnas no deseadas (“NumdeMunicipios”, “ZonasdeBaño”, “PuntosdeMuestreo”) con la función *select()*. Por otro lado, modificaremos las columnas que aparecen con “.x”, recuperando su nombre original seleccionando el primer valor no nulo.

```
grafTablaFinal <- ggplot(tablaend, aes(x = Cantidad, y = EsperanzaDeVida, size = Presupuesto, color = C
  geom_point() +
  labs(title = "Relación entre Cantidad, Esperanza de Vida, Presupuesto y Calidad ", x = "Cantidad
    y = "Esperanza de Vida",
    size = "Presupuesto",
    color = "CalidadAgua") +
  theme_minimal()
grafTablaFinal
```



En relación con el último gráfico, se puede observar la combinación de las cuatro variables tratadas a lo largo de este seminario. Para poder relacionarlas entre ellas, creamos un gráfico de dispersión mediante la función *geom_point()*. El eje de las abscisas representa la Cantidad de agua consumida, el eje de las ordenadas representa la Esperanza de vida, y la Calidad del Agua se encarga de proporcionar el color a los círculos mientras que su tamaño esta determinado por el Presupuesto. Cabe destacar que el único color visible es el morado; esto no quiere decir que el resto de valores no esten representados, sino que se encuentran superpuestos.

Como conclusión final, podemos ver que la relación entre presupuesto y la cantidad de agua consumida si que es notoria. No obstante, se observa que la esperanza de vida no se ve afectada por el presupuesto. Podemos comprobar que los extremos en la variable de la cantidad de agua si que perjudican a la esperanza de vida; aquellos que consumen mucha agua o muy poca tienen una esperanza de vida relativamente inferior a los que consumen una cantidad intermedia. Por otro lado, al superponerse los puntos de calidad de agua, se evidencia que realmente no es una variable que afecte directamente a nuestro estudio. Como conclusión, la esperanza de vida tiende a ser mayor si nuestro consumo de agua no excede ningun límite, permitiendo asi que aquellas Comunidades Autónomas que invierten una gran cantidad en el consumo de agua puedan reducir dicha inversión.

6. Conclusión

A lo largo de este seminario, hemos relacionado las distintas variables con el fin de comprobar si realmente afectan o no a la esperanza de vida. Como conclusión, la esperanza de vida tiende a ser mayor si nuestro consumo de agua no excede ningun límite, permitiendo asi que aquellas Comunidades Autónomas que invierten una gran cantidad en el consumo de agua puedan reducir dicha inversión.

Anexo I

- Orden de la tabla EsperanzayCantidad de mayor a menor en función de:

La esperanza de vida

```
EsperanzaVidaMayorMenor <- EsperanzayCantidad[order(-EsperanzayCantidad$EsperanzaDeVida), ]
EsperanzaVidaMayorMenor
```

```
## # A tibble: 16 x 4
## # Groups:   ComunidadAutonoma [16]
##   Anio ComunidadAutonoma EsperanzaDeVida Cantidad
##   <dbl> <chr>                <dbl>    <dbl>
## 1 2020 BALEARS, ILLES      29.3    42033.
## 2 2020 GALICIA             29.3    71149.
## 3 2020 CANARIAS            29.1    79923.
## 4 2020 NAVARRA, COMUNIDAD FORAL DE 29.0    22551.
## 5 2020 CANTABRIA           28.9    20367.
## 6 2020 PAÍS VASCO          28.8    52635.
## 7 2020 RIOJA, LA           28.6     9448.
## 8 2020 COMUNITAT VALENCIANA 28.5   188854.
## 9 2020 CASTILLA Y LEÓN     28.5    73272.
## 10 2020 ARAGÓN             28.4    38689.
## 11 2020 MURCIA, REGIÓN DE   28.4     57000.
## 12 2020 CATALUÑA           28.3   239767.
## 13 2020 ANDALUCÍA          28.0    262867.
## 14 2020 MADRID, COMUNIDAD DE 28.0    216232.
## 15 2020 EXTREMADURA        27.9    28064.
## 16 2020 CASTILLA - LA MANCHA 27.5    60920.
```

La cantidad de agua consumida

```
CantidadAguaMayorMenor <- EsperanzayCantidad[order(-EsperanzayCantidad$Cantidad), ]
CantidadAguaMayorMenor
```

```
## # A tibble: 16 x 4
## # Groups:   ComunidadAutonoma [16]
##   Anio ComunidadAutonoma EsperanzaDeVida Cantidad
##   <dbl> <chr>                <dbl>    <dbl>
## 1 2020 ANDALUCÍA          28.0    262867.
## 2 2020 CATALUÑA           28.3   239767.
## 3 2020 MADRID, COMUNIDAD DE 28.0    216232.
## 4 2020 COMUNITAT VALENCIANA 28.5   188854.
## 5 2020 CANARIAS            29.1    79923.
## 6 2020 CASTILLA Y LEÓN     28.5    73272.
## 7 2020 GALICIA             29.3    71149.
## 8 2020 CASTILLA - LA MANCHA 27.5    60920.
## 9 2020 MURCIA, REGIÓN DE   28.4     57000.
## 10 2020 PAÍS VASCO          28.8    52635.
## 11 2020 BALEARS, ILLES      29.3    42033.
## 12 2020 ARAGÓN             28.4    38689.
## 13 2020 EXTREMADURA        27.9    28064.
```

```
## 14 2020 NAVARRA, COMUNIDAD FORAL DE      29.0  22551.
## 15 2020 CANTABRIA                        28.9  20367.
## 16 2020 RIOJA, LA                       28.6   9448.
```

- Orden de la tabla Cantidad y Presupuestos en función de Cantidad:

```
CantidadyPresupuestos<- tablaCantidadDeAguaFinal %>%
  left_join(x=., y=tablaPresupuestosFinal, by=c("Anio","ComunidadAutonoma")) %>%
  arrange(desc(Cantidad)) %>%
  drop_na()
CantidadyPresupuestos
```

```
## # A tibble: 16 x 4
## # Groups:   Anio [1]
##   Anio ComunidadAutonoma      Cantidad Presupuesto
##   <dbl> <chr>                <dbl>         <int>
## 1 2020 ANDALUCÍA          262867         43361
## 2 2020 CATALUÑA          239767.         30222
## 3 2020 MADRID, COMUNIDAD DE 216232.         80532
## 4 2020 COMUNITAT VALENCIANA 188854.         39274
## 5 2020 CANARIAS           79923.          4015
## 6 2020 CASTILLA Y LEÓN      73272.          8083
## 7 2020 GALICIA            71149           2594
## 8 2020 CASTILLA - LA MANCHA 60920.           2397
## 9 2020 MURCIA, REGIÓN DE    57000.           3170
## 10 2020 PAÍS VASCO          52635.           3001
## 11 2020 BALEARS, ILLES      42033.          14514
## 12 2020 ARAGÓN             38689.            691
## 13 2020 EXTREMADURA        28064.           2263
## 14 2020 NAVARRA, COMUNIDAD FORAL DE 22551.           5704
## 15 2020 CANTABRIA          20367.            223
## 16 2020 RIOJA, LA          9448.             304
```

- Porcentajes de cada tipo de agua por Comunidades Autónomas:

```
tablaCalidadDeAguaFinalnueva <- tablaCalidadDeAgua %>%
  mutate_at(vars(colNumericas), as.integer) %>%
  select(ComunidadAutonoma, Aguas2, Aguas1, Aguas0) %>%
  mutate(total_aguas=(Aguas2+Aguas1+Aguas0)) %>%
  mutate(Agua2_porcentaje=(Aguas2/total_aguas)*100) %>%
  mutate(Agua1_porcentaje=(Aguas1/total_aguas)*100) %>%
  mutate(Agua0_porcentaje=(Aguas0/total_aguas)*100)

tablaCalidadDeAguaFinalnueva
```

```
##   ComunidadAutonoma Aguas2 Aguas1 Aguas0 total_aguas
## 1 ANDALUCÍA          3      36      27         66
## 2 ARAGÓN             3       7       1         11
## 3 ASTURIAS           0       0       1          1
## 4 CASTILLA - LA MANCHA 24       7      12         43
## 5 CASTILLA Y LEÓN     0       1       1          2
## 6 CATALUÑA           3       8       0         11
```

## 7	EXTREMADURA	0	0	0	0
## 8	GALICIA	10	45	13	68
## 9	MADRID, COMUNIDAD DE	0	2	5	7
## 10	MURCIA, REGIÓN DE	0	0	3	3
## 11	NAVARRA, COMUNIDAD FORAL DE	4	5	2	11
## 12	RIOJA, LA	0	1	0	1
## 13	COMUNITAT VALENCIANA	0	1	1	2
##	Agua2_porcentaje	Agua1_porcentaje	Agua0_porcentaje		
## 1	4.545455	54.54545	40.909091		
## 2	27.272727	63.63636	9.090909		
## 3	0.000000	0.00000	100.000000		
## 4	55.813953	16.27907	27.906977		
## 5	0.000000	50.00000	50.000000		
## 6	27.272727	72.72727	0.000000		
## 7	NaN	NaN	NaN		
## 8	14.705882	66.17647	19.117647		
## 9	0.000000	28.57143	71.428571		
## 10	0.000000	0.00000	100.000000		
## 11	36.363636	45.45455	18.181818		
## 12	0.000000	100.00000	0.000000		
## 13	0.000000	50.00000	50.000000		

Referencias

datos.gob.es. (2023, 22 noviembre). Esperanza de vida por comunidad autónoma, según sexo, edad y nivel educativo. IDB (Identificador API: 37664) - Conjunto de datos. <https://datos.gob.es/es/catalogo/ea0010587-esperanza-de-vida-por-comunidad-autonoma-segun-sexo-edad-y-nivel-educativo-idb-identificador-api-37664>

datos.gob.es. (2022, 6 julio). Distribución de agua registrada por comunidades y ciudades autónomas, grupos de usuarios e importe y periodo. (Identificador API: /T26/P067/P01/serie/L0/01004.Px) - Conjunto de datos. <https://datos.gob.es/es/catalogo/ea0010587-distribucion-de-agua-registrada-por-comunidades-y-ciudades-autonomas-grupos-de-usuarios-e-importe-y-periodo-identificador-api-t26-p067-p01-serie-l0-01004-px>

INE - Instituto Nacional de Estadística. (s. f.). Distribución de agua registrada por comunidades y ciudades autónomas, grupos de usuarios e importe y periodo. INE. <https://www.ine.es/jaxi/Tabla.htm?tpx=53448&L=0>